# Geometrical Method of Asymptotic Conditional Inference Based on the Subset Parameters

Bo-Cheng Wei and Chih-Ling Tsai

University of Minnesota
School of Statistics

Technical Report No. 417

April 1983

University of Minnesota
School of Statistics
Department of Applied Statistics
St. Paul, Minnesota 55108

*The first author is the visiting scholar from The Nanjing Institute of Technology, The People's Republic of China.

Summary

Given a multiparameter curved exponential family with parameter vector $\mu$ which can be partitioned into a component parameter of interest u, and a component nuisance parameter v, we use differential geometry and Edgeworth expansion approach to derive the asymptotic conditional distribution, expectation and variance of an efficient estimator $\hat{u}$ conditioned on an efficient estimator $\hat{v}$. The asymptotic conditional variance of $\hat{u}$ conditioned on an efficient estimator $\hat{v}$ and an ancillary statistic is also derived. If the nuisance parameter v doesn't exist, then the results are exactly the same as given by Amari (1982b).

KEY WORDS: Ancillary Statistics; Conditional Inference; Curved Exponential Family; Curvature; Differential Geometry in Statistics; Edgeworth Expansion; Non-linear Model; QR-decomposition.

## 1. Introduction

Amari (1982b) derived the asymptotic conditional expectation and the asymptotic conditional variance of an efficient estimator $\hat{\mu}$ given an ancillary statistic in a multiparameter curved exponential family. Often the underlying distribution depends not only on a set of parameters u which are of interest, but also on a set of nuisance parameters v. For instance, we may wish to make inferences about the mean u of a normal population with unknown variance v. In the Bayesian approach, inference about u is completely determined by the posterior distribution of u, obtained by "integrating out" the nuisance parameter v from the joint posterior distribution of u and v. In calculating such probabilities, we must have a posterior distribution for v. If no such information on v can be obtained, inference on u can be made based on a sufficient statistic for u.

The traditional conditionality principle specifies that if the minimal sufficient statistic T contains a component a (called an ancillary statistic) whose distribution is independent of $\mu = (u,v)$, then inference about $\mu$ should be based only on the conditional distribution of T given a. Amari constructed a differential geometry theory approach for this type of conditional inference. In this paper, we propose a differential geometry method in obtaining the asymptotic conditional distribution of an efficient estimator $\hat{u}$, given an efficient estimator $\hat{v}$ of the nuisance parameter v in the case of multiparameter curved exponential family. The exponential curvature of a model will be shown to play a fundamental role in the asymptotic theory. Furthermore, the asymptotic conditional variance of $\hat{u}$ given $\hat{v}$ and the ancillary statistics are also obtained. Finally, the asymptotic conditional variance of $\hat{u}$ given $\hat{v}$ is derived for the multiparameter nonlinear model and logistic regression model.

Amari (1982) set an example of constructing a differential-geometrical framework in statistics. The present paper will follow this structure and most of the notation used in Amari's paper.

Denote the set of the distributions of exponential family S by density functions

$$(1.1)\qquad p(x,\theta) = c(X)\exp\{\theta^T X-\psi(\theta)\}$$

where $X = (X_1,\ldots,X_n)^T$ is a random vector in the sample space $x$, $\theta = (\theta^1,\ldots,\theta^n)^T$ is a vector parameter specifying the distributions S with respect to some given measure $m(\cdot)$. We always assume that the necessary regularity conditions are satisfied (see e.g., Barndorff-Nielson, 1980). The set of distributions forms an n-dimensional Riemannian manifold. Its Riemannian metric tensor $g_{ij}(\theta)$ in the $\theta$-coordinate system at $\theta$ is given by the Fisher information matrix as follows

$$(1.2)\qquad g_{ij}(\theta) = E(\partial_i\ell\cdot\partial_j\ell) = \partial_i\partial_j\psi(\theta)$$

where $\ell$ and $\partial_i$ are abbreviations of $\ell(x,\theta) = \log p(x,\theta)$ and $\partial/\partial\theta^i$ respectively. The inverse of matrix $g_{ij}$ is $g^{ij}$. A one-parameter family of affine connections is given by

$$(1.3)\qquad \overset{\alpha}{\Gamma}_{ijk}(\theta) = \frac{1-\alpha}{2}\,T_{ijk}(\theta)$$

where

$$(1.4)\qquad T_{ijk}(\theta) = E(\partial_i\ell\cdot\partial_j\ell\cdot\partial_k\ell) = \partial_i\partial_j\partial_k\psi(\theta)\ .$$

In the tangent space $T_\theta$ of S at $\theta$, $\partial_i\ell(i=1,\ldots,n)$ are n natural basis vectors under the θ-coordinate system. The inner product of two vectors $X = (X^i)$ and $Y = (Y^i)$ in the tangent space $T_\theta$ at θ is given by

$$(1.5) \qquad \langle X, Y \rangle = g_{ij}(\theta)X^iY^j$$

where Einstein's summation convention is used as in the rest of this paper. If the inner product of X and Y is zero, X and Y are orthogonal. A covariance derivative of $X^i \varepsilon T_\theta$ with respect to α-connection is given by

$$(1.6) \qquad \overset{\alpha}{\nabla}_k X^i \equiv \frac{DX^i}{d\theta^k} = \frac{\partial X^i}{\partial \theta^k} + \overset{\alpha}{\Gamma}{}^i_{kj}\, X^j$$

If $\theta^k = \theta^k(\mu)$ $\mu = (\mu^1,\ldots,\mu^m)$, denote

$$(1.7) \qquad \overset{\alpha}{\nabla}_a X^i \equiv \frac{DX^i}{d\mu^a} = \frac{DX^i}{d\theta^k}\frac{\partial\theta^k}{\partial\mu^a} = B^k_a \overset{\alpha}{\nabla}_k X^i \qquad (B^k_a = \frac{\partial\theta^k}{\partial\mu_a})$$

We use notation $\nabla_a X^i$ for α=1 and $\overset{m}{\nabla}_a X^i$ for α= -1 in the present paper.

There are some advantages in using the expectation parameter η in statistics (Amari, 1982a). Where the expectation of $X_i$ is given by

$$(1.8) \qquad EX_i = \eta_i(\theta) = \partial_i\psi(\theta)$$

the mapping (1.8) from θ to η is one-to-one. η can also be used as a coordinate system for S. Any vector X of the sample space x can be treated as a vector in the η-coordinate system since $EX = \eta$.

Denote the set of distributions of curved exponential family M by density functions

$$(1.9) \qquad p(X,\theta(\mu)) = c(X)\exp\{\theta(\mu)^TX-\psi(\theta(\mu))\}$$

where $\mu=(\mu^1,\ldots,\mu^m)^T$ is a vector parameter specifying M. θ(μ) and η(μ) are continuously twice differentiable vector functions of μ. M forms an m-dimension submanifold embedded in S. In the tangent subspace $T_\mu$ of M at $\mu$, $B_A^i(A=1,\dots,m)$ are m basis vectors of $T_\mu$, where

$$(1.10) \qquad B_A^i(\mu) = \frac{\partial\theta^i(\mu)}{\partial\mu^A} = \partial_A\theta^i$$

the Riemannian metric tensor of M at $\theta(\mu)$ is given by

$$(1.11) \qquad g_{AB}(\mu) = B_A^i B_B^j g_{ij}(\theta(\mu)) .$$

The inverse of $g_{AB}$ is $g^{AB}$. Note that $g_{AC}g^{CB} = \delta_A^B$ where $\delta_A^B$ is Kronecker delta. The $\alpha$-connection of the curved exponential family is given by

$$(1.12) \qquad \overset{\alpha}{\Gamma}_{ABC} = (\partial_A B_B^i)B_C^j g_{ij} + \frac{1}{2}(1-\alpha)T_{ABC}$$

where

$$(1.13) \qquad T_{ABC} = B_A^i B_B^j B_C^k \, T_{ijk} .$$

Note that

$$(1.14) \qquad \overset{\alpha}{\Gamma}_{ABC} = \langle \overset{\alpha}{\nabla}_A B_B^i, B_C^j \rangle$$

where $\overset{\alpha}{H}{}^i_{AB} = \overset{\alpha}{\nabla}_A B_B^i$ is called the $\alpha$-curvature.

<u>The notes of index rule</u>. In the present paper, we use notation $\omega = (\mu',a)$, $\mu' = (u',v')$, $\mu = (u,v)$. The indices are used as follows

$\alpha,\beta,\gamma,$ ..... run from 1 to n for $\omega$.

A,B,C, ..... run from 1 to m for $\mu,\mu'$.

a,b,c, ..... run from 1 to k for $u,u'$.

p,q,r, ..... run from k+1 to m for $v,v'$.

$\kappa,\lambda,\delta,$ ..... run from m+1 to n for a .

The tensor notation is used for matrices since it can be easily generalized for multi-index arithmetic. The super-index denotes the number of row and sub-index denotes the number of column.

## 2. Subset Parameters of a Curved Exponential Family

Many authors are interested in ancillary statistic and the associated conditional inference (see, e.g., Efron and Hinkley, 1978; Hinkley, 1980; Barndorff-Nielsen, 1980). The ancillary statistic can be used to recover information loss. However, sometimes one might pay more attention to the parameter itself. Suppose $\mu$ is partitioned into two parts: $\mu=(u,v)$, where $u=(u^1,\ldots,u^k)^T$ is the parameter of interest, $V=(v^{k+1},\ldots,v^{k+\ell})^T$, $k+\ell=m$.

By Amari's methods, Riemannian manifold S can be decomposed into two parts at any point $\theta(\mu)$: submanifold M and its orthogonal complement a -- ancillarly space. According to the orthogonality, many inferences can be made. We will rotate the coordinate system $\mu=(u,v)$ to $\mu'=(u',v')$ to get an orthonormal basis in the tangent subspace $T_\mu$, so that M can be decomposed into two orthogonal submanifolds. Inference can be made for u' and v'.

Let B be the nxm matrix of $B^i_A$, $i=1,\ldots,n$; $A=1,\ldots,m$. We can form the QR decomposition of B proposed by Bates and Watts (1980)

$$(2.1) \qquad B = QR \qquad \text{or} \qquad B^i_A = Q^i_C R^C_A$$

where Q is an nxm matrix with orthogonal column vectors. That means

$$(2.2) \qquad Q^i_A Q^j_B g_{ij} = \delta_{AB}$$

R is an mxm upper triangular matrix. Note that the QR decomposition used in the present paper is slightly different from the ordinary QR decomposition since the inner product here is based on (1.5). So ordinary QR decomposition program cannot be used for computing. But the procedures are almost the same.

Transform the coordinate system $\mu$ to $\mu'$

(2.3) $\mu' = R\mu$ or $\mu'^A = R^A_C \mu^C$

(2.4) $\mu = L\mu'$ or $\mu^A = L^A_C \mu'^C$

where $L = R^{-1}$

$\mu'$ is partitioned into two parts: $\mu' = (u', v')$ corresponding u and v.

By (2.3) and (2.4), the partitioned equations are given by

(2.5) $u'^a = R^a_c u^c + R^a_r v^r$

$v'^p = R^p_r v^r$

(2.6) $u^a = L^a_c u'^c + L^a_r v'^r$

$v^p = L^p_r v'^r$

where $L^A_B = \begin{pmatrix} L^a_b & L^a_q \\ L^p_b & L^p_q \end{pmatrix}$ $R^A_B = \begin{pmatrix} R^a_b & R^a_q \\ R^p_b & R^p_q \end{pmatrix}$

and

a,b run from 1 to k and p,q run from k+1 to m,

(2.7) $L^q_b = R^q_b = 0$ $(1 \le b \le k \ , \quad k+1 \le q \le m)$

Note that (1.10) - (1.14) hold for $\mu'$ coordinate system by adding "'" for related quantities.

After QR decomposition and transformation, the basis of the tangent subspace $T_{\mu'}$ of M at $\mu'$ becomes orthonormal. In fact by (1.11), the metric tensor $g'_{AB}$ of M with respect to coordinate $\mu'$ can be represented by

(2.8) $g'_{AB} = B'^i_A B'^j_B g_{ij}$

By (2.1) and (2.4),

(2.9) $$B'^{\,i}_{A} = \frac{\partial\theta^i}{\partial\mu^C}\,\frac{\partial\mu^C}{\partial\mu'^A} = B^i_C L^C_A = Q^i_D R^D_C L^C_A = Q^i_D \delta^D_A = Q^i_A$$

(2.2) shows that

(2.10) $$g'_{AB} = \delta_{AB}$$

Obviously, $g'^{AB} = \delta^{AB}$. Therefore, when we lower or raise any index of a tensor by multiplying the metric tensor $g'_{AB}$ or its inverse $g'^{AB}$ in $\mu'$-coordinate system, there is no numerical change for that tensor. For example, the value of $T'^{ABC}$ is equal to the value of $T'_{ABC}$.

Since tangent vectors $B'^{\,C}_{A}$ (A=1,...,m) are orthonormal, the tangent subspace $T_{u'}$ spanned by $B'^{\,i}_{a} = \partial\theta^i/\partial u'^a$ (a=1,...,k) is orthogonal to the tangent subspace $T_{v'}$ spanned by $B'^{\,i}_{p} = \partial\theta^i/\partial v'^p$ (p=k+1,...,m). These two tangent subspaces correspond to two certain submanifolds $M_{u'}$ and $M_{v'}$ at $\mu'$. We can study parameter $u'$ and $v'$ instead of $u$ and $v$, then come back by (2.5). Note that the upper triangle matrices L and R give us advantages from (2.7).

Since the transformation matrices L and R and the metric tensors $g_{AB}$ and $g^{AB}$ relate the $\mu$-coordinates with the $\mu'$-coordinates, they are important in discussing the behavior of $u'$ and $v'$. The following formulas are useful:

(2.11) $$L^C_A L^D_B g_{CD} = \delta_{AB}$$

(2.12) $$R^C_A R^D_B \delta_{CD} = g_{AB}$$

(2.13) $$R^A_C R^B_D g^{CD} = \delta^{AB}$$

(2.14) $$L^A_C L^B_D \delta^{CD} = g^{AB}$$

(2.11) comes from (2.8), (2.9) and (1.11), in fact

$$\delta_{AB} = g'_{AB} = B^i_C L^C_A B^j_D L^D_B g_{ij} = L^C_A L^D_B g_{CD}$$

By multiplying inverse of L in (2.11), (2.12) can be proved

$$R^A_E R^B_F \delta_{AB} = R^A_E L^C_A R^B_F L^D_B g_{CD} = \delta^C_E \delta^D_F g_{CD} = g_{EF}$$

then taking inverse of (2.11) and (2.12), (2.13) and (2.14) can be obtained.

Corresponding to u and v, we partition related matrices $g_{AB}$ and $g^{AB}$

$$g_{AB} = \begin{pmatrix} g_{ab} & g_{aq} \\ g_{pb} & g_{pq} \end{pmatrix} \qquad g^{AB} = \begin{pmatrix} g^{ab} & g^{aq} \\ g^{pb} & g^{pq} \end{pmatrix}$$

where a.b run from 1 to k and p.q run from k+1 to m. The equation (2.11)-(2.14) are not necessarily true for index a,b,c,..., and p,q,r,.... In fact by (2.7), the partitioned equations for (2.13) have the following form:

(2.15) $$g^{ab} = L^a_c L^b_d \delta^{cd} + L^a_r L^b_s \delta^{rs}$$

(2.16) $$g^{ap} = L^a_r L^p_s \delta^{rs}$$

(2.17) $$g^{pq} = L^p_r L^q_s \delta^{rs}$$

By (2.15)-(2.17), more useful formulas can be obtained

(2.18) $$\bar{g}_{pq} = R^r_p R^s_q \delta_{rs} \qquad \text{where } \bar{g}_{pq} = (g^{pq})^{-1}$$

(2.19) $$g^{ab} - g^{ap}\bar{g}_{pq}g^{qb} = L^a_c L^b_d \delta^{cd}$$

(2.20) $$\delta^{pq} = R^p_r R^q_s g^{rs}$$

## 3. Conditional Distribution

Suppose we sample $X_1,\ldots,X_N$ independently from the curved exponential family with density $p(X,\theta(\mu))$ at $\mu$ and $\overline{X} = N^{-1}\sum_{i=1}^{N} X_i$ is a sufficient statistic. Let $\hat{\mu}$ be a consistent, first order efficient estimator based on $\overline{X}$ and let $a(\mu)$ be the associated ancillary family (Amari 1981). Obviously, $\hat{\mu}' = R\hat{\mu}$ is also a consistent, first order efficient estimator and $a(\mu')$ is its associated ancillary family. First, we concentrate on the estimator $\hat{\mu}'$ of the parameter $\mu'=R\mu$. Suppose the local coordinate system $(\mu',a)$ in some neighborhood around $\eta(\mu')$ has been con- constructed. The a-coordinate corresponds to ancillary space. The coordinate of the point of M is $(\mu',0)$. Put $\omega=(\mu',a)=(u',v',a)$ (see section 1 for index rule). When dealing with $\overline{X}$ as a point of $\eta$-coordinate system, $\overline{X}=\eta(\hat{\mu}',\hat{a})$ and $\hat{\omega}=(\hat{\mu}',\hat{a})$ form a sufficient statistic (Amari, 1981).

Note that (1.10)-(1.14) hold for $\omega$-coordinate system by replacing $\alpha,\beta,\gamma$ for A,B,C. For example, the metric tensor in the $\omega$-coordinate system is written as

$$g_{\alpha\beta} = B_\alpha^i B_\beta^j g_{ij} \qquad \text{where } B_\alpha^i = \partial\theta^i/\partial\omega^\alpha \ (\alpha=1,\ldots,n)$$

$g_{\alpha\beta} = g'_{AB} = \delta_{AB}$ if $\alpha,\beta$ run from 1 to m, $g_{\alpha\beta}=g_{\kappa\lambda}$ if $\alpha,\beta$ run from m+1 to n. Otherwise, $g_{\alpha\beta}=g_{A\kappa}=0$.

In order to obtain the Edgeworth expansion of the estimators, $\hat{\omega}$ has to be standardized to $\tilde{\omega}$:

$$(3.1) \qquad \tilde{\omega} = \sqrt{N}\,(\hat{\omega}-E\hat{\omega})$$

By Amari's paper (1981, 1982b)

$$(3.2) \qquad E\hat{\omega}^\alpha = \omega^\alpha + b^\alpha/N$$

$$(3.3) \qquad b^\alpha = -1/2\ C_{\beta\gamma\delta}g^{\alpha\delta}g^{\beta\gamma} = -1/2\ C^\alpha_{\beta\gamma}g^{\beta\gamma}$$

where

(3.4) $$C_{\beta\gamma\delta} = \overset{m}{\Gamma}_{\beta\gamma\delta} = \partial_\beta(\partial_\gamma \eta_i)\cdot(\partial_\delta \eta_j) g^{ij}$$

When $\beta,\gamma,\delta \le m$, the symbol $b'^A$ is the bias of $\hat{\mu}'^A$

(3.5) $$b'^A = -\frac{1}{2}\overset{m}{\Gamma}{}'^A_{BC}\delta^{BC} - \frac{1}{2}\overset{m}{H}{}'^A_{\kappa\lambda} g^{\kappa\lambda}$$

where $\overset{m}{H}{}'^A_{\kappa\lambda}$ is a curvature tensor of the ancillary space. It vanishes when $\hat{\mu}$ is an ML estimator. (3.5) shows that the bias of ML estimator is independent of the a-coordinate. Define:

(3.6) $$\bar{\omega} = \sqrt{N}(\hat{\omega}-\omega)$$

then by (3.1)

(3.7) $$\tilde{\omega}^\alpha = \bar{\omega}^\alpha - b^\alpha/\sqrt{N} \quad .$$

It means that if the tolerance error is up to $O(1/\sqrt{N})$, $\tilde{\omega}$ can be replaced by $\bar{\omega}$ without loss. $\bar{\omega}$ is useful to eliminate the bias term.

Amari gave Edgeworth expansions for $\tilde{\omega}$, $\tilde{\mu}'$ and $\tilde{a}$. We only need the expansion of $\tilde{\mu}' = (\tilde{u}', \tilde{v}')$

(3.8) $$p(\tilde{\mu}') = \psi(\tilde{\mu}')\{1 + \frac{1}{6\sqrt{N}} K'_{ABC}\tilde{H}'^{ABC}(\tilde{\mu}') + O(1/N)\}.$$

where $$\psi(\tilde{\mu}') = (1/\sqrt{2\pi})^m \exp\{-\frac{1}{2}\delta_{AB}\tilde{\mu}'^A\tilde{\mu}'^B\}$$

$$K'_{ABC} = T'_{ABC} - C'_{ABC} - C'_{BCA} - C'_{CAB}$$

$\tilde{H}'^{ABC}(\tilde{\mu}')$ are multidimensional Hermite polynomials of $\tilde{\mu}'$ (Amari, 1981).

(3.9) $$\tilde{H}'^{ABC}(\tilde{\mu}') = \tilde{\mu}'^A\tilde{\mu}'^B\tilde{\mu}'^C - \delta^{AB}\tilde{\mu}'^C - \delta^{BC}\tilde{\mu}'^A - \delta^{CA}\tilde{\mu}'^B \quad .$$

By integrating (3.8) with respect to $\tilde{u}'$ and using

$$(3.10)\qquad K'_{ABC}\tilde{H}'^{ABC} = K'_{abc}\tilde{H}'^{abc} + 3K'_{abp}\tilde{H}'^{ab}\tilde{H}'^{p} + 3K'_{apq}\tilde{H}'^{a}\tilde{H}'^{pq} + K'_{pqr}\tilde{H}'^{pqr}$$

where $\tilde{H}'^{ab} = \tilde{u}'^{a}\tilde{u}'^{b} - \delta^{ab}$ $\qquad \tilde{H}'^{a} = \tilde{u}'^{a}$

The expansion of $\tilde{v}'$ can be obtained by

$$(3.11)\qquad p(\tilde{v}') = \psi(\tilde{v}')\{1 + \frac{1}{6\sqrt{N}} K'_{pgr}\tilde{H}'^{pgr}(\tilde{v}') + O(1/N)\}.$$

In order to return $\tilde{\mu}$ and $\tilde{v}$ from (3.8) and (3.11), we use (2.3), (2.5) and (2.12). The expansion $p(\tilde{\mu})$ and $p(\tilde{v})$ are given by

$$(3.12)\qquad p(\tilde{\mu}) = c\ \exp\{-\frac{1}{2} g_{AB}\tilde{\mu}^{A}\tilde{\mu}^{B}\}\{1 + \frac{1}{6\sqrt{N}} K'_{ABC}\tilde{H}'^{ABC}(\mu'\to\mu) + O(1/N)\},$$

$$(3.13)\qquad p(\tilde{v}) = c\ \exp\{-\frac{1}{2} \delta_{pq} R^{p}_{r}R^{q}_{s}\tilde{v}^{r}\tilde{v}^{s}\}\{1 + \frac{1}{6\sqrt{N}} K'_{pqr}\tilde{H}'^{pqr}(\tilde{v}'\to\tilde{v}) + O(1/N)\}.$$

where we always denote integral constant by the same notation c without loss of generality. $\tilde{H}'^{ABC}(\tilde{\mu}'\to\tilde{\mu})$ are the abbreviation of substituting $\mu'=R\mu$ for (3.9).

(3.12) shows that distribution of $\tilde{\mu}$ is asymptotically normal with covariance $g^{AB}$. By (2.18), $R^{p}_{r}R^{q}_{s}\delta_{pq}$ equals inverse of $g^{rs}$. So $p(\tilde{v})$ is asymptotically normal with covariance $g^{rs}$. It is the marginal distribution of $\tilde{\mu}$ in the asymptotic sense. Therefore, the first theorem can be obtained by using (3.10) and (2.20).

<u>Theorem 1</u>. The conditional distribution of $\tilde{u}$ given by $\tilde{v}$ is given by

$$(3.14)\qquad p(\tilde{u}|\tilde{v}) = Q(\tilde{u},\tilde{v})\{1 + \frac{1}{6\sqrt{N}}[K'_{abc}\tilde{H}'^{abc}(\tilde{u}'\to\tilde{u}) + 3K'_{abp}R^{p}_{r}\tilde{H}'^{ab}(\tilde{u}'\to\tilde{u})\tilde{H}^{r}(\tilde{v})$$

$$+ 3K'_{apq}R^{p}_{r}R^{q}_{s}\tilde{H}'^{a}(\tilde{u}'\to\tilde{u})\tilde{H}^{rs}(\tilde{v})] + O(1/N)\},$$

where $Q(\tilde{u},\tilde{v}) = c\ \exp\{-\frac{1}{2}(\tilde{u}^{a}-L^{a}_{p}R^{p}_{q}\tilde{v}^{q})R^{c}_{a}R^{d}_{b}\delta_{cd}(\tilde{u}^{b}-L^{b}_{r}R^{r}_{s}\tilde{v}^{s})$ .

$p(\tilde{u}|\tilde{v})$ is asymptotically normal with

(3.15) $\quad E(\tilde{u}^a|\tilde{v}) = L^a_p R^p_q \tilde{v}^q + O(1/\sqrt{N}) = L^a_p R^p_q \overline{v}^q + O(1/\sqrt{N})$

(3.16) $\quad \mathrm{Var}(\tilde{u}^a, \tilde{u}^b|\tilde{v}) = L^a_c L^b_d \delta^{cd} + O(1/\sqrt{N})$ .

Remark 1. (2.15) - (2.19) show that

(3.17) $\quad L^a_p R^p_q = g^{ap}\overline{g}_{pq}$

(3.18) $\quad (R^c_a R^d_b \delta_{cd})^{-1} = L^a_c L^b_d \delta^{cd} = g^{ab} - g^{ap}\,\overline{g}_{pq} g^{qb}$

They match the ordinary conditional expectations and covariances in the multi-normal case.

Remark 2. $\tilde{u}$ and $\tilde{v}$ can be replaced by $\overline{u}$ and $\overline{v}$ in the right hand of (3.14) except first term $Q(\tilde{u},\tilde{v})$ to eliminate the bias term without effect on the order of magnitude of the error. By (3.15) and (3.16) the expressions of expectation and covariance with error $O(1/\sqrt{N})$ are independent of the a-coordinate.

Remark 3. By Amari's paper (1982,b) and (1.14)

$$K'_{abc} = -3 < \overset{\alpha}{\nabla}{}'_a B'^{\,i}_b ,\ B'^{\,j}_c > = -3\overset{\alpha}{\Gamma}{}'_{abc} \qquad (\alpha = -\tfrac{1}{3})$$

$< \overset{\alpha}{\nabla}_a B'^{\,i}_b ,\ B'^{\,j}_c >$ is the projection of the $\alpha$-curvature of submanifold $M_{u'}$ onto tangent subspace $T_{u'}$ (Amari, 1982a). It is not a tensor so $K'_{abc}$ depends on coordinate system.

$K'_{abp} = < \nabla'_a B'^{\,i}_b ,\ B'^{\,j}_p > = H'_{abp}$ is the intrinsic curvature tensor of $M_{u'}$. So $K'_{abp}$ is an invariant.

$K'_{apq} = - < \overset{m}{\nabla}{}'_p B'^{\,i}_q ,\ B'^{\,j}_a > = -\overset{m}{H}{}'_{pqa}$ is the intrisic curvature tensor of $M_{v'}$. $K'_{apq}$ is also an invariant.

All those quantities do not depend on ancillary space.

## 4. Conditional Expectation and Covariance

It might be hard to calculate the conditional expectation and covariance more precisely by using (3.14), since the expression involves some complex calculations. By (2.6), the calculation can be done by using the conditional distribution $p(\tilde{u}'|\tilde{v}')$.

By (3.10), distribution of $\tilde{\mu}'$ can be rewritten as

$$(4.1) \qquad p(\tilde{u}',\tilde{v}') = \psi(\tilde{u}')\psi(\tilde{v}')\{1 + \frac{1}{6\sqrt{N}} K'_{pqr}\tilde{H}'^{pqr}(\tilde{v}') + O(1/N)\}$$

$$\cdot\{1 + \frac{1}{6\sqrt{N}} [K'_{abc}\tilde{H}'^{abc}(\tilde{u}') + 3K'_{abp}\tilde{H}'^{ab}(\tilde{u}')\tilde{H}'^{p}(\tilde{v}') + 3K'_{apq}\tilde{H}'^{a}(\tilde{u}')$$

$$\cdot\tilde{H}'^{pq}(\tilde{v}')] + O(1/N)\} .$$

The conditional distribution of $\tilde{u}'$ given $\tilde{v}'$ is obtained by

$$(4.2) \qquad p(\tilde{u}'|\tilde{v}') = \Psi(\tilde{u}')\{1 + \frac{1}{6\sqrt{N}} [K'_{abc}\tilde{H}'^{abc}(\tilde{u}') + 3K'_{abp}\tilde{H}'^{ab}(\tilde{u}')\tilde{H}'^{p}(\tilde{v}')$$

$$+ 3K'_{apq}\tilde{H}'^{a}(\tilde{u}')\tilde{H}'^{pq}(\tilde{v}')] + O(1/N)\} .$$

Taking (2.6) into account, it is easy to compute conditional expectation and covariance of $\tilde{u}$ by (4.2) with the aid of the orthogonality of the multidimensional Hermite polynomials.

$$E(\tilde{u}^a|\tilde{v}) = E(L^a_c\tilde{u}'^c + L^a_r\tilde{v}'^r|\tilde{v}') = L^a_r R^r_s\tilde{v}^s + L^a_c E\{\tilde{H}'^c(\tilde{u}')|\tilde{v}'\} \quad .$$

Note that

$$\int \tilde{H}'^{c}(\tilde{u}')p(\tilde{u}'|\tilde{v}')d\tilde{u}' = \frac{1}{2\sqrt{N}} K'_{apq}\delta^{ac}\tilde{H}'^{pq}(\tilde{v}') + O(1/N)$$

Replacing $\tilde{v}$ by $\bar{v}$ for $\tilde{H}'^{pq}(\tilde{v}'\to\tilde{v})$ and taking (2.5) and (2.20) into account, we obtain

Theorem 2. The conditional expectation of an efficient estimator $\hat{u}$ given $\hat{v}$ is obtained by

$$\text{(4.3)} \qquad E(\hat{u}^a|\hat{v}) = u^a + \frac{b^a}{N} + L^a_r R^r_s(\Delta\hat{v}^s - \frac{b^s}{N}) - \frac{1}{2}\sum_{c=1}^{k} L^a_c R^p_r R^q_s H'^{m}_{pqc}$$

$$\cdot\,(\Delta\hat{v}^r\Delta\hat{v}^s - g^{rs}/N) + O(N^{-3/2})$$

where $\Delta\hat{v}^r = \hat{v}^r - v^r$.

The fourth term of the right hand of (4.3) is in terms of relative curvature. It is added to adjust the ordinary conditional expectation of multi-normal case (first three terms).

In order to compute covariance, the following formula is needed:

$$\text{(4.4)} \qquad \mathrm{Var}(\tilde{u}^a,\tilde{u}^b|\tilde{v}) = E(\tilde{u}^a\tilde{u}^b|\tilde{v}) - E(\tilde{u}^a|\tilde{v})E(\tilde{u}^b|\tilde{v}) \quad .$$

Note that $\tilde{u}'^a\tilde{u}'^b = \tilde{H}'^{ab} + \delta^{ab}$ and

$$\int \tilde{H}'^{ab}(\tilde{u}')p(\tilde{u}'|\tilde{v}')d\tilde{u}' = \frac{\tilde{H}'^{p}(\tilde{v}')}{2\sqrt{N}} K'_{cdp}(\delta^{ac}\delta^{bd} + \delta^{ad}\delta^{bc})$$

After some simple calculation, we obtain

Theorem 3. The conditional covariance of an efficient estimator $\hat{u}$ given $\hat{v}$ is obtained by

$$(4.5)\quad \mathrm{Var}(\hat{u}^a,\hat{u}^b|\hat{v}) = \frac{1}{N}\{L^a_c L^b_d \delta^{cd} + \sum_{c=1}^{k}\sum_{d=1}^{k} L^a_c L^b_d R^p_r H'_{cdp}\,\Delta\hat{v}^r\} + O(N^{-2})$$

Similarly, the conditional distribution, expectation and covariance of $\hat{u}$ given $\hat{v}$ and $\hat{a}$ can be also obtained. For example, we have

Theorem 4. The conditional covariance of an efficient estimator $\hat{u}$ given $\hat{v}$ and $\hat{a}$ is obtained by

$$(4.6)\quad \mathrm{Var}(\hat{u}^a,\hat{u}^b|\hat{v},\hat{a}) = \frac{1}{N}\{L^a_c L^b_d \delta^{cd} + \sum_{c=1}^{k}\sum_{d=1}^{k} L^a_c L^b_d (H'_{cdp} R^p_r\,\Delta\hat{v}^r + H'_{cd\kappa}\hat{a}^\kappa)\} + O(N^{-2})$$

This result is similar to (4.5). If $\hat{v}$ is not given, that means k=m, $H'_{cdp}=0$. (4.6) reduces to Amari's result (1982,b), see appendix for details.

## 5. Examples in Non-Linear Model and Logistic Regression Model

Example 1: Non-Linear Model

Drapper and Smith (1981) defined a model as non-linear in the parameters if that model cannot be written as

$$(5.1)\qquad \theta(X,\mu) = \mu^j g_j(X)$$

where $g_j(X)$ is any function of the independent variable X. Let $y_{ij}$ (i=1,...,n, j=1,...,N) collected at corresponding experimental settings $x_i$ (i=1,...,n). It is assumed that the relationship between the responses and the experimental settings can be represented by an equation of the form

$$(5.2)\qquad y_{ij} = \theta(X_i,\mu) + \varepsilon_{ij}$$

where $\mu = (u',\ldots,u^k, v^{k+1},\ldots,v^m)$ is a set of unknown parameters and $\varepsilon_{ij}$ is an additive error component with normal distributed, where

$$E(\varepsilon_{ij}) = 0, \; i=1,\ldots,n, \; j=1,\ldots,N \quad ,$$

and $E(\varepsilon_{ij}\,\varepsilon_{\ell j}) = \delta_{i\ell}$, $i,\ell=1,\ldots,n$, $j=1,\ldots,N$. The probability density of $y_\ell = (y_{1\ell},\ldots,y_{n\ell})^T$ is

$$(5.3) \qquad P(y_\ell,\theta) = c\,\exp\{-\tfrac{1}{2}(y_\ell-\theta)^T(y_\ell-\theta)\}.$$

The set of such probability density function $P(y,\theta)$ which belongs to exponential family forms a n-dimensional manifold S. Hence, the metric tensor $g_{ij}$ and $\alpha$-connection $\Gamma^{\alpha}_{ijk}$ can be calculated by equation (1.2) and (1.3).

$$(5.4) \qquad g_{ij}(\theta) = E(\partial_i\ell\,\partial_j\ell) = \delta_{ij}$$

$$(5.5) \qquad \overset{\alpha}{\Gamma}_{ijk}(\theta) = \frac{1-\alpha}{2}\,T_{ijk} = \frac{1-\alpha}{2}\,E(\partial_i\ell\,\partial_j\ell\,\partial_k\ell) = 0 \;.$$

Because $\theta_i$ is also a function of parameter $\mu$, $P(y,\theta(\mu))$ is a m-dimensional curved exponential family of a large n-parameter exponential family.

The metric tensor and $\alpha$-connection over the m-submanifold are calculated from equations (1.11) and (1.12) as

$$(5.6) \qquad g_{AB}(\mu) = B^i_A B^j_B g_{ij}(\theta(\mu))$$

$$= B^i_A B^j_B \delta_{ij} \qquad \text{and}$$

$$(5.7) \qquad \overset{\alpha}{\Gamma}_{ABC} = (\partial_A B^i_B) B^j_C g_{ij} + \frac{1}{2}(1-\alpha)\;T_{ABC}$$

$$= (\partial_A B^i_E) B^j_C \delta_{ij}$$

because $T_{ABC} = B^i_A B^j_B B^k_C\;T_{ijk} = 0$

We apply equation (3.5), equation (5.7), remark 3 and theorm 4 to the multiparameter non-linear model case and get the following result.

(5.8) $\quad \mathrm{Var}(\hat{u}|\hat{v}) = \frac{1}{N}\{\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} - \tilde{L}^T[(\Delta\hat{v}')^T][A^{N1}_{\cdot\cdot}]\tilde{L}\} + O(\frac{1}{N^2})$

where $LL^T = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$, $\Sigma_{11}$, $\Sigma_{12}$ and $\Sigma_{22}$ are k by k, k by (m-k) and (m-k) by (m-k) submatrices of $LL^T$, respectively; $\Sigma_{11} - \Sigma_{12}\,\Sigma_{22}^{-1}\,\Sigma_{21}$ is the conditional variance of $\hat{u}$ given $\hat{v}$ for the non-linear model with linear approximation; $\tilde{L}$ is the first k by k submatrix of L; $A^{N1}_{\cdot\cdot}$ contains the first k by k submatrices of the last (m-k) components in the parameter effects array $A^T_{\cdot\cdot}$ which was defined by Bates and Watts (1980) and [·] [·] is the bracket multiplication which was also defined by them.

Example 2: Logistic Regression Model

Given a sample of n independent binominal response $y_i \sim B(n_i,p_i)$, the log likelihood function for the sample is the sum of individual likelihood contributions:

$$\ell(\theta,y) = \sum_{i=1}^{n} \ell(\theta^i,y_i)$$

$$= y_i\theta^i - a(\theta) + b(y)$$

where $b(y) = \sum_{i=1}^{n} \log\begin{pmatrix} n_i \\ y_i \end{pmatrix}$, $a(\theta) = n_i \log(1+e^{\theta^i})$

and $\quad \theta^i = \log \frac{P_i}{1-P_i}$ .

The logistic regression specifies the relationship $\theta$ = logist (P) = X $\mu$ where $P = (P_1,\ldots,P_n)^T$, $\theta = (\theta^1,\ldots,\theta^n)^T$

$\mu = (\mu^1,\ldots,\mu^m)^T$, $X = (X_1,\ldots,X_m)$ and $X_a = (X_a^1,\ldots,X_a^n)^T$, $a=1,\ldots,m$.

Therefore, the set of densities of the logistic regression model belongs to the curved exponential family.

The metric tensor $g_{ij}$ and $\alpha$-connection $\overset{\alpha}{\Gamma}_{ijk}$ over the n-dimensional manifold can be calculated as $g_{ij}(\theta) = E(\partial_i \ell\, \partial_j \ell) = \delta_{ij} \dfrac{n_j \exp(\theta^j)}{(1+\exp(\theta^i))^2}$

and $\overset{\alpha}{\Gamma}_{ijk}(\theta) = \frac{1-\alpha}{2} E(\partial_i \ell\, \partial_j \ell\, \partial_k \ell)$

$$= \frac{1-\alpha}{2} \delta_{ij}\delta_{jk} \frac{n_j(\exp(\theta^j))(1-\exp(\theta^k))}{(1 + \exp(\theta^i))^3}$$

The matric tensor $g_{ab}$, $\alpha$-connection $\overset{\alpha}{\Gamma}_{abc}$ and -curvature $\overset{\alpha}{H}{}^i_{ab}$ over the m-dimensional submanifold can be calculated as

$$g_{ab} = X_a^i X_b^j g_{ij} \quad ,$$

$$\overset{\alpha}{\Gamma}_{abc} = X_a^i X_b^j X_c^k \overset{\alpha}{\Gamma}_{ijk}$$

and

$$\overset{\alpha}{H}{}^i_{ab} = \overset{\alpha}{\Gamma}{}^i_{jk}\, X_a^i X_b^k$$

Now we apply remark 3 and theorm 4 to the logistic regression model with N identically independent replications at each experimental points x and get the following result.

$$H'_{abp} = H'^{i}_{ab} B'^{j}_{p}\, g_{ij} = 0$$

and

$$\mathrm{Var}(\hat{u}^a,\hat{u}^b|\hat{v}) = \frac{1}{N} \{L_c^a L_d^b \delta_d^c\} + O(\frac{1}{N^2}) \ .$$

Therefore, the conditional variance of $\hat{u}$ given $\hat{v}$ in the logistic regression model is independent of exponential curvature.

Acknowledgement

We would like to express our sincere gratitude to Professor D. V. Hinkley, who has given us a great deal of useful advice in the preparation of this paper. We also thank Professor R. D. Cook for his helpful discussion and correspondence on this topic.

## APPENDIX

The proof of theorem 4:

The Edgeworth expansion of the density function of $\tilde{\omega}$ is given by

$$(A.1) \qquad p(\tilde{\omega}) = \psi(\tilde{u}')\psi(\tilde{v}')\phi(\tilde{a})\{1 + \frac{1}{6\sqrt{N}} K_{\alpha\beta\gamma} \tilde{H}^{\alpha\beta\gamma}(\tilde{\omega}) + O(\tfrac{1}{N})\}$$

where $\phi(\tilde{a}) = c \exp\{-\tfrac{1}{2} g_{\kappa\lambda}\tilde{a}^{\kappa}\tilde{a}^{\lambda}\}$ .

Since $g_{A\kappa} = o$ $(1 \le A \le m,\ m < \kappa \le n)$ and $g_{ap} = o$ $(1 \le a \le k,\ k < p \le m)$. $K_{\alpha\beta\gamma}\tilde{H}^{\alpha\beta\gamma}$ can be decomposed by

$$(A.2) \qquad K_{\alpha\beta\gamma}\tilde{H}^{\alpha\beta\gamma} = K'_{abc}\tilde{H}'^{abc} + 3K'_{abp}\tilde{H}'^{ab}\tilde{H}'^{p} + 3K'_{apq}\tilde{H}'^{a}\tilde{H}'^{pq} + K'_{pqr}\tilde{H}'^{pqr}$$
$$+ 3K'_{ab\kappa}\tilde{H}'^{ab}\tilde{H}^{\kappa} + 6K'_{ap\kappa}\tilde{H}'^{a}\tilde{H}'^{p}\tilde{H}^{\kappa} + 3K'_{pq\kappa}\tilde{H}'^{pq}\tilde{H}^{\kappa}$$
$$+ 3K'_{a\kappa\lambda}\tilde{H}'^{a}\tilde{H}^{\kappa\lambda} + 3K'_{p\kappa\lambda}\tilde{H}'^{p}\tilde{H}^{\kappa\lambda} + K_{\kappa\lambda\delta}\tilde{H}^{\kappa\lambda\delta}$$

(Amari, 1981). Integrating (A.1) with respect to $\tilde{u}'$, the expansion of density $p(\tilde{v}',\tilde{a})$ can be written as

$$(A.3) \qquad p(\tilde{v}',\tilde{a}) = \psi(\tilde{v}')\phi(\tilde{a})\ \{1 + \frac{1}{6\sqrt{N}}[K'_{pqr}\tilde{H}'^{pqr} + 3K'_{pq\kappa}\tilde{H}'^{pq}\tilde{H}^{\kappa} + 3K'_{p\kappa\lambda}\tilde{H}'^{p}\tilde{H}^{\kappa\lambda}$$
$$+ K_{\kappa\lambda\delta}\tilde{H}^{\kappa\lambda\delta}] + O(\tfrac{1}{N})\}.$$

By (A.1), (A.2) and (A.3), the expansion of the conditional distribution of $\tilde{u}'$ is obtained by

$$p(\tilde{u}'|\tilde{v}',\tilde{a}) = \psi(\tilde{u}')\{1 + \frac{1}{6\sqrt{N}}[K'_{abc}\tilde{H}'^{abc} + 3K'_{abp}\tilde{H}'^{ab}\tilde{H}'^{p} + 3K'_{apq}\tilde{H}'^{a}\tilde{H}'^{pq}$$
$$+ 3K'_{ab\kappa}\tilde{H}'^{ab}\tilde{H}^{\kappa} + 6K'_{ap\kappa}\tilde{H}'^{a}\tilde{H}'^{p}\tilde{H}^{\kappa} + 3K'_{\alpha\kappa\lambda}\tilde{H}'^{a}\tilde{H}^{\kappa\lambda}] + O(\tfrac{1}{N})\}.$$

By the orthogonality of multidimensional Hermite polynomials, it is easy to calculate

(A.5) $$E(\tilde{u}'^{a}|\tilde{v}',\tilde{a}) = \frac{\delta^{ac}}{2\sqrt{N}}[K'_{cpq}\tilde{H}'^{pq} + 2K'_{cp\kappa}\tilde{H}'^{p}\tilde{H}^{\kappa} + K'_{c\kappa\lambda}\tilde{H}^{\kappa\lambda}] + O(\tfrac{1}{N})$$

(A.6) $$E(\tilde{u}'^{a}\tilde{u}'^{b}|\tilde{v}',\tilde{a}) = \delta^{ab} + \frac{\delta^{ac}\delta^{bd}}{\sqrt{N}}[K'_{cdp}\tilde{H}'^{p} + K'_{cd\kappa}\tilde{H}^{\kappa}] + O(\tfrac{1}{N}) \quad .$$

The theorem 4 can be followed by (A.5), (A.6) and (4.4). If $\hat{v}$ is not given, then k=m, (4.6) reduces to

(A.7) $$\mathrm{Var}(\hat{\mu}^{A},\hat{\mu}^{B}|\hat{a}) = \frac{1}{N}\{L^{A}_{C}L^{B}_{D}\delta^{CD} + \sum_{C=1}^{m}\sum_{D=1}^{m} L^{A}_{C}L^{B}_{D}H'_{CD\kappa}\hat{a}^{\kappa}\} + O(N^{-2}) \ ,$$

where

$$L^{A}_{C}L^{B}_{D}\delta^{CD} = g^{AB}$$

$$H'_{CD\kappa} = \langle \nabla'_{C}B'^{i}_{D},\ B^{j}_{\kappa}\rangle = L^{E}_{C}L^{F}_{D}\langle \nabla_{E}B^{i}_{F},\ B^{j}_{\kappa}\rangle = L^{E}_{C}L^{F}_{D}H_{EF\kappa}$$

noting that

$$\sum_{C=1}^{m}\sum_{D=1}^{m} L^{A}_{C}L^{B}_{D}H'_{CD\kappa}\hat{a}^{\kappa} = \delta^{CE}\delta^{DF}L^{A}_{C}L^{B}_{D}H'_{EF\kappa}\hat{a}^{\kappa} \quad \text{and}$$

$$\delta^{CD} = R^{C}_{E}R^{D}_{F}g^{EF} \quad .$$

(A.7) reduces to

$$\mathrm{Var}(\hat{\mu}^{A},\hat{\mu}^{B}|\hat{a}) = \frac{1}{N}(g^{AB} + H^{AB}_{\kappa}\hat{a}^{\kappa}) + O(N^{-2}) \ .$$

This is the same as Amari's result (1982,b).

## REFERENCES

AMARI, S. (1982,a). Differential geometry of curved exponential families—curvatures and information loss. Ann. Statist. 10, 357-385.

AMARI, S. (1982,b). Geometrical theory of asymptotic ancillarity and conditional inference. Biometrika 69, 1-17.

AMARI, S. and KUMON, M. (1981). Differential geometry of Edgeworth expansions in curved exponential family. Technical Report, METR 81-7, University of Tokyo.

BARNDORFF-NIELSEN, O. (1980). Conditional resolutions. Biometrika 67, 293-310.

BATES, D.M. and WATTS, D.G. (1980). Relative curvature measures of nonlinearity. J. Roy. Statist. Soc. B 42, 1-25.

DRAPER, N. and SMITY, H. (1981). Applied Regression Analysis, 2nd ed., John Wiley & Sons, Inc.

EFRON, B. and HINKLEY, D.V. (1978). Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information (with discussion), Biometrika 65, 457-87.

HINKLEY, D.V. (1980). Likelihood as approximate pivotal distribution. Biometrika 67, 287-92.

Bo-Cheng Wei
Department of Basic Sciences
Nanjing Institute of Technology
People's Republic of China
(After September 30, 1983)

Chih-Ling Tsai
Department of Applied Statistics
and Operations Research
New York University
100 Trinity Place
New York, NY 10006
(After August 30, 1983)